CS 134: Networks

Prof. Yaron Singer Spring 2017

1 Overview

In this lecture, we begin with some context about the Web, and then look further back into the history of information networks that led up to the Web, and finally check out one powerful method of "node endorsement in such a network, namely **PageRank**.

2 The World Wide Web

Definition. *A **path** from a node A to a node B in a directed graph is a sequence of nodes, beginning with A and ending with B, with the property that each consecutive pair of nodes in the sequence is connected by an edge pointing in the forward direction.*

Definition. *A directed graph is **strongly connected** if there is a path from every node to every other node.*

Definition. *A **strongly connected component** (SCC) in a directed graph is a subset of the nodes such that*

- *every node in the subset has a path to every other;*
- *the subset is not part of some larger set with the property that every node can reach every other*

3 PageRank

ALG 1

input: Graph G with n nodes, number of steps t

1. Assign all nodes the same initial PageRank value, $1/n$;
2. For t iterations:
   Update all PageRank values according to the *Basic PageRank Update Rule*;
   End for

return: The PageRank value of each node

Definition. ***Basic PageRank Update Rule**: Each page divides its current PageRank equally across its outgoing links and passes these equal shares to the pages it points to. (If a page has no outgoing links, it passes all its current PageRank to itself.) Each page updates its new PageRank to be the sum of the shares it receives.*

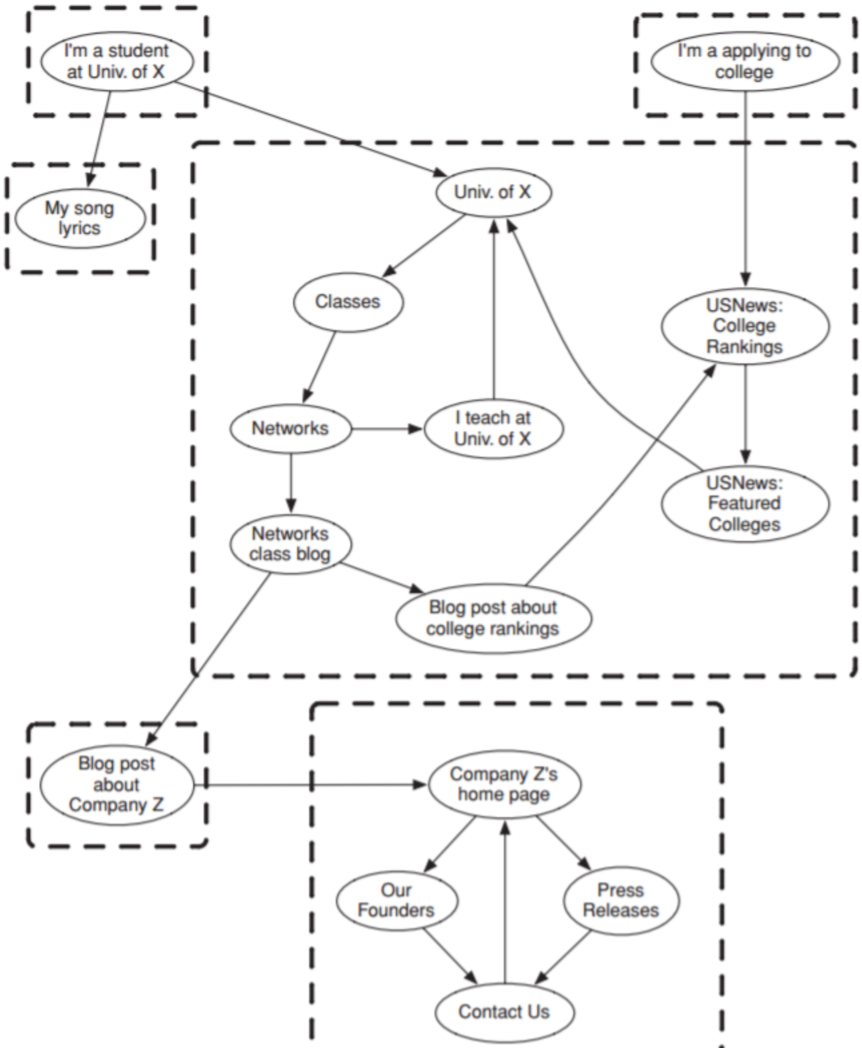


Figure 1: The directed graph in Figure 1 with its strongly connected components identified. (*Kleinberg 13.3*).

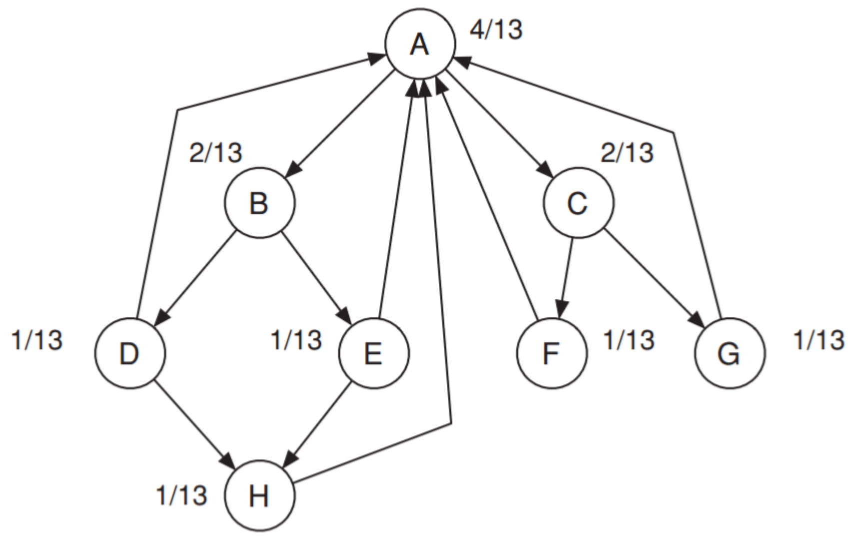


Figure 2: A collection of eight pages with their equilibrium PageRank values: page A has the largest PageRank, followed by pages B and C (which collect endorsements from A). (*Kleinberg 14.3*).

Let M be an n by n matrix, where n is the number of nodes in the networks and M_{ij} is the share of page i's PageRank that j should get in one update step. Thus, we have:

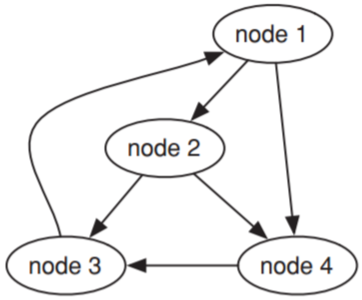
$$M_{ij} = \begin{cases} 1/d(i) & \text{if } i \text{ points to } j \\ 1 & \text{if } d(i) = 0 and i = j \\ 0 & \text{otherwise} \end{cases}$$

where $d(i)$ is the outgoing degree of i. If we represent the PageRanks of all nodes at step t using a vector $r^{(t)}$, then to calculate an entry in $r^{(t)}$ we have:

$$r_j^{(t)} = M_{1j} \cdot r_1^{(t)} + \cdots + M_{nj} \cdot r_n^{(t)}.$$

$$r^{(t)} = M^T \cdot r^{(t-1)}$$

$$r^{(t)} = (M^T)^t \begin{bmatrix} 1/n \\ \cdots \\ 1/n \end{bmatrix}.$$



$$\begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Figure 3: The flow of PageRank under the Basic PageRank Update Rule can be represented using a "transition" matrix M. (*Kleinberg 14.6*).

4 Scaled PageRank

Definition. ***Scaled PageRank Update Rule***: *First apply the Basic PageRank Update Rule. Then scale down all PageRank values by a factor of s. We divide the residual $1 - s$ units of PageRank equally over all nodes, giving $\frac{1-s}{n}$ to each.*

If we are applying the Scaled PageRank Update Rule to our transition matrix M, we calculate new values in our new "scaled" matrix $\hat{M}$ as follows:

$$\hat{M}_{ij} = sM_{ij} + \frac{1-s}{n}$$

Then the following equations all hold true:

$$r_j^{(t)} = \hat{M}_{1j} \cdot r_1^{(t)} + \cdots + \hat{M}_{nj} \cdot r_n^{(t)}.$$

$$r^{(t)} = \hat{M}^T \cdot r^{(t-1)}$$

$$r^{(t)} = (\hat{M}^T)^t \begin{bmatrix} 1/n \\ \cdots \\ 1/n \end{bmatrix}.$$

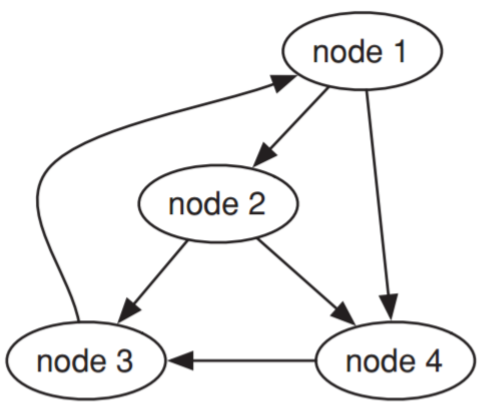
Theorem. *(Perron's Theorem) For any matrix P whose entries are all positive:*

1. *P has a real eigenvalue $\lambda_1 > 0$ such that $\lambda_1 > |\lambda'|$ for all other eigenvalues λ'.*

2. *There is an eigenvector x with positive real terms corresponding to the largest eigenvalue λ_1, and x is unique up to multiplication by a constant.*

3. *If $\lambda_1 = 1$, then for any starting vector $y \neq \vec{0}$ with non-negative coordinates, where $\vec{0}$ is a vector whose terms all equal zero, the sequence of vectors $P^k y$ converges to a vector in the direction of x as k goes to infinity.*

Using Perron's Theorem, we find that x, the eigenvector corresponding to the largest real eigenvalue of $\hat{M}$ (which will always be 1 since each of the rows in $\hat{M}$ sums to 1), corresponds to the set of equilibrium PageRank values for Scaled PageRank.



$$\begin{bmatrix} .05 & .45 & .05 & .45 \\ .05 & .05 & .45 & .45 \\ .85 & .05 & .05 & .05 \\ .05 & .05 & .85 & .05 \end{bmatrix}$$

Figure 4: A "transition" matrix $\hat{M}$ under the Scaled PageRank Update Rule (with scaling factor $s = 0.8$) and the corresponding network (*Kleinberg 14.6*).